


Chapter 42

Detection of Shotgun Surgery and Message Chain Code Smells using Machine Learning Techniques

Thirupathi Guggulothu

 <https://orcid.org/0000-0002-9081-2816>
University of Hyderabad, Hyderabad, India

Salman Abdul Moiz

University of Hyderabad, Hyderabad, India

ABSTRACT

Code smell is an inherent property of software that results in design problems which makes the software hard to extend, understand, and maintain. In the literature, several tools are used to detect code smell that are informally defined or subjective in nature due to varying results of the code smell. To resolve this, machine learning (ML) techniques are proposed and learn to distinguish the characteristics of smelly and non-smelly code elements (classes or methods). However, the dataset constructed by the ML techniques are based on the tools and manually validated code smell samples. In this article, instead of using tools and manual validation, the authors considered detection rules for identifying the smell then applied unsupervised learning for validation to construct two smell datasets. Then, applied classification algorithms are used on the datasets to detect the code smells. The researchers found that all algorithms have achieved high performance in terms of accuracy, F-measure and area under ROC, yet the tree-based classifiers are performing better than other classifiers.

INTRODUCTION

Code smells or bad code smells refers to an anomaly in the source code that may result in deeper problems which makes software difficult to understand, evolve, and maintain. According to (Booch, 2006) smell is a kind of structure in the code that shows a violation of basic design principles such as Abstraction, Hierarchy, Encapsulation, Modularity, and Modifiability. Even if the design principles are known to the developers due to inexperience, the competition that is in the market and deadline pressure are lead-

DOI: 10.4018/978-1-6684-6291-1.ch042

ing to violation of these principles. Fowler et al. (Fowler, 1999) have defined 22 informal code smells which are removed through refactoring techniques. These techniques are used to enhance the internal structure of the code without varying the external behaviour and to improve the quality of the software. The (Opdyke, 1992) authors have defined 72 refactoring techniques.

There are various methods and tools available in the literature to detect the code smells. Each technique and tool produces different (Fontana, 2012). Bowes et al. (Bowes, 2013), compared two code smell detection tools on message chaining and shown disparity of results between them. The three main reasons for varying results are: 1) The code smells can be subjectively interpreted by the developers, and hence detected in different ways. 2) Agreement between the detectors is low, i.e., different tools or rules detect a different type of smell for different code elements. 3) The threshold value for identifying the smell can vary for the detectors.

To address the above limitations, in particular the subjective nature, Fontana et al. (Fontana, 2016) proposed a machine learning (ML) technique to detect four code smells (Long Method, Data Class, Feature Envy, Large Class) with the help of 32 classification techniques. The authors have built 4 datasets, one for each smell. These datasets have been prepared based on the tools and manual labelling process. Tools are used to identify whether the code elements (instances) are smelly or not. But the tools may produce some false positive instances so, the authors manually validated the instances to avoid the biasness. In this paper, instead of using tools and manual validation, the authors have prepared two new method level code smell datasets of Fowler et al. (Fowler, 1999) from the literature; based on the detection rules and unsupervised learning i.e., clustering to validate the instances as smelly or not.

In the proposed work, an attempt is made to detect two code smells namely Shotgun surgery and Message chaining with supervised learning techniques. It is an application of machine learning (ML) classification approach used for code smell detection. It uses known data to determine how the new instances should be classified into binary classification i.e., based on the metrics used for a particular method, the ML approach helps in classifying a method to be prone to code smell or not. In this paper, the dataset instances are methods of 74 heterogeneous java systems. The metrics of object-oriented systems have been computed on method instances, which are the features or attributes of the dataset. For each smell, one dataset is prepared by using detection rules from the literature (Ferme, 2013). The researchers applied a random stratified sampling on the method instances to balance the datasets. Sample instances of the dataset are validated through unsupervised learning and added to the training dataset. Then applied some known classification algorithms on the trained datasets to detect the code smells, by using 10-fold cross validation method. To evaluate those algorithms, standard metric measures such as F-score, accuracy and the area under the ROC are used. The experimented algorithms have achieved high performance in both the smells.

The paper is been arranged as follows; The second section, introduces a work related to detection of code smells; The third section, defines two proposed approaches of code smell detections; The fourth section, detecting code smells using ML approach; The fifth section, presents experimental results; The sixth section, presents the code smell detection rules; and the final section, gives conclusion and future directions.

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/detection-of-shotgun-surgery-and-message-chain-code-smells-using-machine-learning-techniques/307484

Related Content

Drug Prediction in Healthcare Using Big Data and Machine Learning

Mamoon Rashid, Vishal Goyal, Shabir Ahmad Parahand Harjeet Singh (2022). *Research Anthology on Machine Learning Techniques, Methods, and Applications* (pp. 1062-1071).

www.irma-international.org/chapter/drug-prediction-in-healthcare-using-big-data-and-machine-learning/307498

Analysis of Gravitation-Based Optimization Algorithms for Clustering and Classification

Sajad Ahmad Rather and P. Shanthi Bala (2020). *Handbook of Research on Big Data Clustering and Machine Learning* (pp. 74-99).

www.irma-international.org/chapter/analysis-of-gravitation-based-optimization-algorithms-for-clustering-and-classification/241371

Early Warning System Framework Proposal, Based on Big Data Environment

Goran Klepac, Robert Kopal and Leo Mrcic (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 35-66).

www.irma-international.org/article/early-warning-system-framework-proposal-based-on-big-data-environment/233889

Implementation of Machine Learning-Aided Speech Analysis for Speaker Accent Identification Applied to Audio Forensics

Vijayalakshmi G. V. Mahesh, Alex Noel Joseph Rajand Ruban Nersisson (2022). *Aiding Forensic Investigation Through Deep Learning and Machine Learning Frameworks* (pp. 174-194).

www.irma-international.org/chapter/implementation-of-machine-learning-aided-speech-analysis-for-speaker-accent-identification-applied-to-audio-forensics/309781

Palmprint And Dorsal Hand Vein Multi-Modal Biometric Fusion Using Deep Learning

Norah Abdullah Al-johani and Lamiaa A. Elrefaei (2020). *International Journal of Artificial Intelligence and Machine Learning* (pp. 18-42).

www.irma-international.org/article/palmprint-and-dorsal-hand-vein-multi-modal-biometric-fusion-using-deep-learning/257270